



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### A common framework for language comprehension and language production?

**Citation for published version:**

Hartsuiker, RJ & Pickering, MJ 2001, 'A common framework for language comprehension and language production?', *Behavioral and Brain Sciences*, vol. 24, no. 5, pp. 887-+.  
<[http://journals.cambridge.org/download.php?file=%2FBBS%2FBBS24\\_05%2FS0140525X01000103a.pdf&code=aaa3b7457514c8f5d990e75cf703ee45](http://journals.cambridge.org/download.php?file=%2FBBS%2FBBS24_05%2FS0140525X01000103a.pdf&code=aaa3b7457514c8f5d990e75cf703ee45)>

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Behavioral and Brain Sciences

**Publisher Rights Statement:**

©Hartsuiker, R. J., & Pickering, M. J. (2001). A common framework for language comprehension and language production?. *Behavioral and Brain Sciences*, 24(5), 887-+

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



the functional connections among its products. Thus, perception and action should not be studied in isolation; most likely perceiving-acting is not a decomposable system.

The main assumption of the target article is a logical corollary of this broader claim and, as such, we agree with it. However, the Theory of Event Coding (TEC) does not fulfill its promise, and it makes the very mistake it is meant to correct.

We will focus on the authors' brief, dismissive comments about the motor theory of speech perception. The authors provide two reasons for their dismissal. First, they do not wish to claim that their account applies to language. Second, in the opinions of Jusczyk (1986) and Levelt (1989), empirical support for the motor theory is weak. We address what the authors prefer not to because we consider the reasons for their dismissal mistaken.

However, we will not reply to the second reason here. Evidence from half a century of research cannot be summarized as tersely as a dismissal can be offered (for a review, see Liberman 1996). As for the first reason, speech is an exquisite example of perceptually guided action, and thus it must be addressed by a theory of perception-action codings. Moreover, precisely because it cannot be reduced to the stimulus/response experimental settings the authors choose to use, it is the right place to look for a better understanding of the couplings between perception and action. And the empirical facts that led to the motor theory reveal the flaws of TEC to which we have alluded.

In the tasks that support TEC, experimenters devise stimuli that can be described by sets of arbitrarily-chosen, arbitrarily-combined features (e.g., a letter is red or green; a rectangle is on the right or left side of a computer screen). In the tasks, some of the features are relevant, and so the participants are encouraged to make use of them to do the task, and they do. However, these sets are not up to the task of constituting percepts or action plans in nature. Proponents of TEC have to answer, among others, the following questions:

- 1) Can such feature sets compose real perception-action codings?
- 2) Are percepts linear combinations of features?
- 3) If they are to refer to the distal world, perception-action codings must be grounded. How are features grounded?

The motor theory of speech perception was designed to answer these questions, which arose when Alvin Liberman and colleagues sought the features that capture our perceptions of syllables (Liberman 1957; Liberman et al. 1967; Liberman & Mattingly 1989).

They started where Hommel and colleagues would have liked them to start. During the fifties, as a part of a project to build a reading machine for the blind, they tried several ways to teach people to perceive syllables as strings of featurally distinct sounds. They devised "acoustic alphabets": acoustic features corresponded to single letters, and syllables were sequences of discrete acoustic letters. After years of effort, the project failed. Speech is not an acoustic alphabet, and people could not perceive acoustic alphabetic sequences at practically useful rates.

This realization provides an answer to our first question: arbitrarily-chosen, arbitrarily-combined features do not correspond to perception-action codings.

The next move of Liberman and colleagues was also aligned with TEC. Perhaps the sounds created by the vocal tract have a special character for which arbitrary sounds cannot substitute. Using the sound spectrograph, Liberman and colleagues searched for acoustic features that, they hoped, would characterize syllables. Again, their hopes were frustrated. Spectrograms, far from clarifying the picture, presented a new puzzle. Depending on the context, the same acoustic feature could specify different phonemes and, conversely, different acoustic features could specify the same phoneme. Acoustic features do not capture the dimensions of the space where syllables live.

We can now answer our second question: percepts are not necessarily linear combinations of acoustic features. But, more importantly, this second failure taught Liberman and colleagues a lesson. Once they rejected features as components of speech per-

cepts, they recognized that there is an invariant among different tokens of the same phoneme; it is in the gestures that produced them. This discovery led to a major revision of their scientific assumptions. Motor competence, not the feature set conceived by the scientist, underlies speech perception. The motor theory was born, and it provided a surprising answer to our third question: speech perception is not grounded in the realm of perception, but in the very place where, according to the common coding of distal events, it should be grounded: in the actions of the vocal tract!

To understand the nature of perception-action codings, a major revision of the conceptual apparatus used to implement TEC is needed, beginning with careful decisions about what should and should not count as an observable. Liberman, and with him Roger Sperry (1952), suggested that the use of two different sets of observables – one for perception and one for action – is misleading; motoric observables capture cognition. Although we do not entirely accept this full motorization of cognition, the hard learned lessons that led to the motor theory tell us that the observables conceived by a theory of disembodied cognition will not work. Here lies the major limitation of TEC. It rejects the unreasonable division between action and perception at the level of distal events, while claiming that "late action" and "early perception" are not necessary to explain cognitive common coding. If not there, among the nuts and bolts of the contact between the nervous system and the physical world, where are we to find the grounding of perception-action distal identity?

## A common framework for language comprehension and language production?

Robert J. Hartsuiker and Martin J. Pickering

Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ  
United Kingdom. [Rob.Hartsuiker@ed.ac.uk](mailto:Rob.Hartsuiker@ed.ac.uk)  
<http://www.psy.ed.ac.uk/~robh/> [Martin.Pickering@ed.ac.uk](mailto:Martin.Pickering@ed.ac.uk)

**Abstract:** Natural language processing involves a tight coupling between action (the production of language) and perception (the comprehension of language). We argue that similar theoretical principles apply to language processing as to action/perception in general. Language production is not driven solely by the speaker's intentions; language comprehension is not only input-driven; production and perception use common representations. We will relate recent findings from our language production lab to the Theory of Event Coding (TEC)'s principle of feature binding.

**Hommel et al.** consider standard approaches studying action and perception in isolation to be flawed. They argue that action does not start solely from an organism's goals and intentions, nor is perception input-driven only. The authors present what they refer to as a set of "metatheoretical principles" (TEC) in which action and perception are mutually dependant processes, sharing a common representational medium (event codes). In this commentary, we will argue that these metatheoretical principles can be and should be extended to the field of natural language processing (psycholinguistics), a field of study that the authors explicitly exclude from their target article (see sect. 2.1.4.1). However, in our opinion, language should not necessarily fall outside the scope of their *theory*, for two reasons. First, language processing in natural dialogue consists of a cycle of tightly coupled events, that consecutively involve action (producing an utterance) and perception (perceiving the interlocutor's utterance). These events are closely related. However, like older research traditions in action or perception, research in language has mainly considered production and comprehension as isolated events, ignoring this tight coupling. Second, **Hommel et al.** argue for a common representational medium for action and perception. We argue that the same is true for linguistic acting and perceiving.

### 1. The fundamental concern of psycholinguistics should be

**dialogue, not monologue.** Current research on language processing tends to view production and comprehension as isolated events. Views on language production can perhaps best be characterized by the subtitle of Levelt's (1989) influential book, *Speaking from intention to articulation*, which clearly reveals the essential ideomotor view that presently dominates production research. Likewise, research on language comprehension mostly studies the reading of isolated words or sentences. This line of research attempts to exclude responses (perhaps most successfully in ERP studies on sentence comprehension, where the lack of the need for a response is regarded as a virtue) or to ignore them (e.g., by tacitly assuming that word naming latencies reflect only perceptual aspects of reading; see Bock 1996). However, such research does not do justice to the fact that language is usually produced in dialogue, where perception/production interactions are of central importance (Pickering & Garrod, submitted). Consider a question-answer pair. It is not only the goals and intentions of the answer-giver that drive the content and form of the utterance: it is also the content and form of the question that is being asked. There are also perhaps less obvious phenomena in dialogue that suggest a tight coupling between production and perception. Almost any statement calls for a particular type of response from the interlocutor; dialogue partners often co-produce utterances (you start a sentence and I finish it); dialogue partners tend to "align" their representations (i.e., they tend to use the same linguistic units at each level of processing); they monitor and correct not only their own speech but also that of the other. Any theory of language use needs to account for this coupling between production and comprehension processes.

**2. Common representations for production and perception.** A fundamental assumption of TEC is that action and perception make use of common codes. Similarly, there is growing consensus in psycholinguistics that common representations are used in language production and language comprehension, although theorists are still arguing whether this is only true for abstract lexical representations (Levelt et al. 1999), or for most linguistic levels (MacKay 1987; Pickering & Garrod, submitted). One piece of evidence for this parity of representations is the finding of syntactic priming from comprehension to production (Branigan et al. 2000). Producing a particular syntactic form (e.g., a passive) increases the likelihood of that form being produced again on a subsequent trial (Bock 1986). However, Branigan et al. showed that *perceiving* a particular sentence type is sufficient to create a priming effect in production, suggesting a common code for perceptual and production codes for syntax. According to Pickering and Garrod, common representations play a crucial role in dialogue, because they allow a pair of speakers to achieve "alignment" (similar representations at each level). This alignment is their main theoretical construct for explaining the fluency of dialogue.

**3. Effects of feature overlap in language processing.** One example of the possible viability of TEC as an explanatory mechanism for linguistic processes comes from the domain of syntactic priming. There is now a large body of evidence from our group that this effect is modulated by the degree of lexical overlap (e.g., Pickering & Branigan 1998): the more words are shared between prime and target, the more syntactic repetition. Why is this? From a TEC perspective, one could consider a sentence an integrated representation binding syntactic information and lexical information. If a particular structure, say a prepositional object dative, is bound to particular words (e.g., *the nun showing an apple to a pirate*) and if a subsequent picture shows one or more of these elements (say, the nun giving an apple to a priest), then it would be relatively costly to say *the nun handing the priest the apple* because it would require undoing the binding between particular words and a particular phrasal configuration. The more words are shared, the costlier the undoing of these bindings would become.

**4. Conclusion.** Current-day psycholinguistics is still too much focussed on sentence-production and comprehension in isolated contexts (monologue) and tends to ignore action/perception interactions, except in some circumstances such as self-monitoring

of speech (Hartsuiker & Kolk 2001; Levelt 1989) or in some approaches that stress the role for production in comprehension (Garrett 2000). We think that psycholinguistics would benefit from the controlled study of language in dialogue, casting findings in theoretical frameworks such as TEC.

## TEC – Some problems and some prospects

Julian Hochberg

Psychology Department, Columbia University, New York, NY 10027.  
hochberg@columbia.edu

**Abstract:** The Theory of Event Coding (TEC) is a significant contribution to the study of purposeful perceptual behavior, and can be made more so by recognizing a major context (the work of Tolman, Liberman, Neisser); some significant problems (tightening predictions and defining distal stimuli); and an extremely important area of potential application (ongoing anticipation and perceptual inquiry, as in reading and movies).

Hommel et al.'s paper on the Theory of Event Coding (TEC) is a significant contribution toward a science of purposeful perception and action, one in which the course of intended actions, like key-pressing, is determined by component action codes that share a common cognitive representation with stimulus feature codes ("red," "round," etc.). Supporting research mainly concerns interactions between stimulus and response features as associatively formed during instructed repetitive practice. The authors stress that the stimuli that most directly determine the event are not proximal but distal, by which they mean the perceived features in represented space. This, they rightly contrast with the common approach in which stimuli and responses are defined in terms of the physically measurable proximal stimulus input and effector output. (This particular use of the word "distal" may prove momentarily confusing to perception psychologists, since it is not a 2D projection vs. 3D layout that is at issue.)

Their approach is not quite as uncommon or isolated as their presentation suggests. Tolman's historic analyses (1932; 1948) of animals' purposive behaviors in, and memories of, their behavioral environments in response to the patterns that offer a "means-end readiness" (since termed an "affordance" by Gibson 1979; see also Lewin's valence-driven events in life space), were instantiated by specific experimental demonstrations, like place learning, in which behavior was far better predicted by measurable variables in distal space than by proximal stimulus measures. This approach was not abandoned, nor is it rare. Research on place learning and its purpose-relevant stimuli is not obsolete (cf. Gallistel 1990), and research into the brain processes of topological learning is cutting edge (cf. Zinyuk et al. 2000). Indeed, closely related "constructivist" approaches in which behavior is explained in terms of events that transpire in an internalized space abound today, although not in connection with Reaction Time (RT) experiments, and with little work until now on the explicit relationship between the goal, the stimulus information, and the action.

What is uncommon about TEC is that it appears to set the stage for systematic co-analyses of such purposive behavior, in which the signal and the action are closely related and commensurate. There are, however, several serious problems to be resolved and opportunities grasped, before we can be as comfortable as we might be with TEC as a framework. Namely, in ascending order of difficulty:

(1) More explicit modeling is needed than has been reported in the target article, in order to reach conclusions that are more definite than the ones the paper typically offers, for example, that such-and-such an outcome is not incompatible with TEC.

(2) Identifying and codifying of feature codes is needed if they are to be applied by more than intuition. The "distal stimulus" needs principled, circumstance-dependent definitions: the authors obviously do *not* mean measures defined only in terms of the